

Deteksi Komentar Spam Menggunakan Algoritma Naive Bayes Berbasis PHP-ML

Rizky Parlika^{1*}, Gayuh Abdi Mahardika², Imam Muslem³, Zakial Vikki⁴, Zulkifli⁵

¹ Program Studi Teknologi Informasi, UPN "Veteran" Jawa Timur, Jl. Rungkut Madya, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur 60294, Indonesia. e-mail: rizkyparlika.if@upnjatim.ac.id

² Program Studi Informatika, UPN "Veteran" Jawa Timur, Jl. Rungkut Madya, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur 60294, Indonesia. e-mail: 22081010067@student.upnjatim.ac.id

³ Program Studi Informatika, Universitas Almuslim, Jl. Tengku Abdurrahman No. 37, Matanggumpangdua, Peusangan, Bireuen, Aceh 24261, Indonesia. e-mail: imamtkj@gmail.com

⁴ Program Studi Informatika, Universitas Islam Kebangsaan Indonesia, Jl. Medan - Banda Aceh, Desa Blang Bladeh, Kecamatan Jeumpa, Kabupaten Bireuen, Aceh 24251, Indonesia. e-mail: zakialviki.mkom@gmail.com

⁵ Program Studi Informatika, Universitas Almuslim, Jl. Tengku Abdurrahman No. 37, Matanggumpangdua, Peusangan, Bireuen, Aceh 24261, Indonesia. e-mail: zulladasicupak@gmail.com

*Corresponding Author:

Rizky Parlika

rizkyparlika.if@upnjatim.ac.id

ORCID: <https://orcid.org/0000-0003-2474-930X>

ARTICLE INFO

Article History:

Received : 26 Mei 2026

Revised : 26 Mei 2026

Accepted : 11 Juni 2026

Keywords:

Spam comment detection, php-ml, naive bayes, machine learning

Kata kunci:

Deteksi komentar spam, php-ml, naive bayes, pembelajaran mesin



This work is licensed under a [Creative Commons Attribution-4.0 International License](https://creativecommons.org/licenses/by/4.0/)

NOVAKOMPUTA

Jurnal Ilmu Komputer dan Informatika
Vol 01 No 02 | Juni 2026

ABSTRACT

[Spam Comment Detection Using PHP-ML Based Naive Bayes Algorithm]

Spam comments are a common issue in online interaction systems such as blogs, forums, and e-commerce platforms. This study designs and implements a spam comment classification system as a RESTful API using Lumen and the PHP-ML library. The system accepts comments through API endpoints and classifies them as spam or not (ham) using the Naive Bayes algorithm. All received comments are stored in a MySQL database and can be manually relabeled to support model retraining. The retraining process is executed via command-line interface, allowing the system to incorporate new data from both the database and external CSV datasets. Evaluation results show that the system can classify comments in real-time with high accuracy and supports model updates without disrupting the main service. This approach is considered effective and adaptable for web applications that require automated and continuous spam detection capabilities.

ABSTRAK

Komentar spam merupakan permasalahan umum dalam sistem interaksi daring seperti blog, forum, dan e-commerce. Penelitian ini merancang dan mengimplementasikan sistem klasifikasi komentar spam berbasis REST API menggunakan Lumen dan pustaka PHP-ML. Sistem menerima komentar dari client aplikasi melalui endpoint API, lalu mengklasifikasikannya sebagai spam atau bukan (ham) menggunakan algoritma Naive Bayes. Komentar yang diterima juga disimpan ke dalam database MySQL dan dapat dilabeli ulang secara manual untuk mendukung retraining model secara berkala. Proses pembelajaran ulang dilakukan melalui command-line interface untuk memanfaatkan data baru dari basis data maupun dataset eksternal. Hasil pengujian menunjukkan bahwa sistem mampu mengklasifikasikan komentar secara real-time dengan tingkat akurasi yang tinggi, serta menyediakan fleksibilitas pembaruan model tanpa mengganggu layanan utama. Sistem ini dinilai efektif dan adaptif untuk diterapkan pada aplikasi web yang

membutuhkan deteksi spam komentar secara otomatis dan berkelanjutan.

PENDAHULUAN

Komentar merupakan salah satu bentuk interaksi paling umum dalam berbagai platform web seperti forum, blog, marketplace, hingga media sosial [1]. Namun, seiring dengan meningkatnya penggunaan fitur komentar, muncul pula tantangan besar berupa penyebaran komentar spam. Komentar spam biasanya berisi promosi tidak relevan, phishing, atau tautan berbahaya yang tidak hanya mengganggu pengguna, tetapi juga berisiko merusak reputasi situs [2]. Penanganan komentar spam menjadi semakin penting dalam menjaga keamanan dan kenyamanan pengguna di ranah digital [3], [4].

Dalam beberapa tahun terakhir, berbagai pendekatan berbasis machine learning telah dikembangkan untuk mendeteksi komentar spam secara otomatis. Salah satu metode yang paling banyak digunakan karena kesederhanaan dan efektivitasnya adalah algoritma Naive Bayes [5]. Metode ini cocok untuk klasifikasi teks karena mampu menangani data berskala besar dengan efisiensi tinggi [6], [7]. Model klasifikasi komentar yang dilatih dengan algoritma ini terbukti dapat mengidentifikasi pola dalam bahasa spam dengan akurasi yang cukup baik [8], [9].

Seiring dengan kebutuhan sistem yang fleksibel dan mudah diintegrasikan dengan berbagai platform, pendekatan berbasis REST API menjadi solusi ideal [10]. REST API memungkinkan layanan klasifikasi dapat diakses secara real-time oleh sistem eksternal, seperti website atau aplikasi mobile, tanpa perlu integrasi langsung dengan sistem machine learning di sisi backend [11]. Pendekatan ini juga memungkinkan pengembangan sistem yang modular dan scalable [12].

Dalam penelitian ini, penulis mengembangkan sistem klasifikasi komentar spam berbasis PHP-ML dan diimplementasikan dalam bentuk REST API menggunakan framework Lumen, yaitu micro-framework dari Laravel yang ringan dan cocok untuk kebutuhan API [13], [14]. Sistem ini dirancang agar tidak hanya melakukan klasifikasi secara instan, tetapi juga menyimpan komentar ke basis data MySQL untuk memungkinkan pelabelan ulang secara manual oleh admin [15], [16]. Pelabelan ulang ini digunakan untuk retraining model secara berkala, sehingga sistem dapat terus belajar dari data baru dan menjadi semakin adaptif [17].

Fitur pembelajaran ulang menjadi salah satu keunggulan utama dalam penelitian ini. Dengan menyimpan semua komentar yang diterima beserta hasil prediksi dan (jika tersedia) label kebenaran, sistem dapat melakukan retraining dengan memanfaatkan data-data baru yang lebih representatif [18]. Hal ini memungkinkan sistem untuk tetap relevan meskipun bahasa dan pola komentar spam terus berkembang [19]. Proses retraining dilakukan melalui perintah CLI khusus, yang memungkinkan pengelola sistem melakukan pembaruan model secara manual maupun terjadwal [20].

Adapun masalah utama yang diangkat dalam penelitian ini adalah: bagaimana merancang sistem klasifikasi komentar spam yang ringan, akurat, dan dapat diakses melalui API, serta mampu memperbarui model klasifikasinya secara dinamis? Untuk menjawab hal tersebut, dibutuhkan kombinasi pendekatan teknologi yang tepat, baik dari sisi algoritma, integrasi API, hingga desain database. Penelitian ini diharapkan mampu memberikan kontribusi nyata terhadap pengembangan sistem deteksi spam yang fleksibel dan mudah diintegrasikan dengan berbagai platform web.

METODE

A. Desain Penelitian

Penelitian ini bersifat terapan (*applied research*), dengan pendekatan rekayasa perangkat lunak. Tujuan utama adalah merancang dan mengembangkan sistem klasifikasi komentar spam berbasis REST API. Sistem dibangun menggunakan framework Lumen, library PHP-ML untuk algoritma Naive Bayes, serta MySQL sebagai basis data. Pengembangan dilakukan secara iteratif, mencakup tahap analisis kebutuhan, perancangan arsitektur sistem, implementasi, pengujian, dan evaluasi.

B. Alat dan bahan

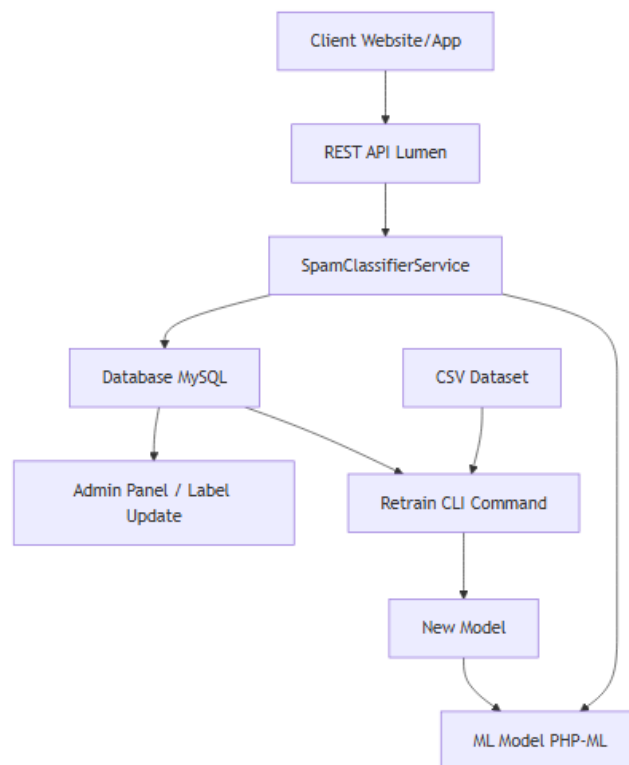
Perangkat keras: laptop dengan prosesor minimal Intel Core i5, RAM 8 GB. Perangkat lunak:

- PHP 8.x, Lumen 8.x, Composer
- PHP-ML (library machine learning untuk PHP)
- MySQL untuk penyimpanan data komentar

- d) Postman untuk pengujian endpoint
- e) Visual Studio Code sebagai IDE
- f) Dataset: *YouTube Spam Collection Dataset* dari Kaggle, yang telah dibersihkan dan diolah menjadi format CSV serta ditambahkan ke dalam database lokal.

C. Arsitektur Sistem

Sistem dikembangkan dengan pendekatan mikroservis. Komponen utama adalah REST API (Lumen), layanan klasifikasi berbasis PHP-ML, dan penyimpanan data komentar di MySQL. Data yang diklasifikasi disimpan untuk keperluan pelabelan ulang dan retraining model. Arsitektur sistem divisualisasikan pada Gambar 2.1.

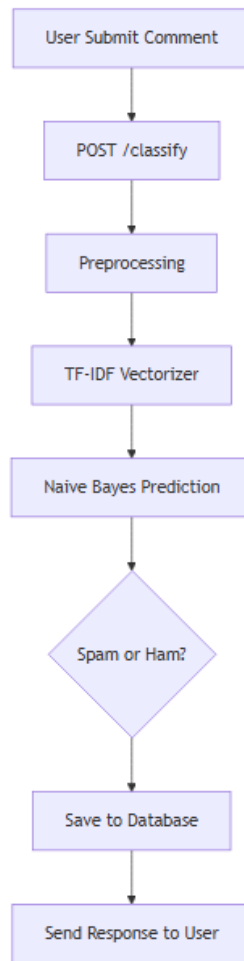


Gambar 1. Diagram Arsitektur Sistem

Diagram ini menunjukkan alur utama antara client, API, service klasifikasi, dan model. Sistem mendukung pembelajaran berkelanjutan melalui retraining model dari data baru yang diperoleh di database dan CSV. Pendekatan seperti ini telah diterapkan pada berbagai sistem klasifikasi berbasis web (Rafique et al., 2021; Ibrahim et al., 2021).

D. Alur Proses Klasifikasi Komentar

Setiap komentar yang dikirim melalui client web akan diproses oleh endpoint POST /api/classify. Sistem kemudian melakukan preprocessing dan klasifikasi menggunakan model Naive Bayes yang telah dilatih. Proses ini digambarkan dalam Gambar 2.2.



Gambar 2. Diagram Alur Klasifikasi Komentar

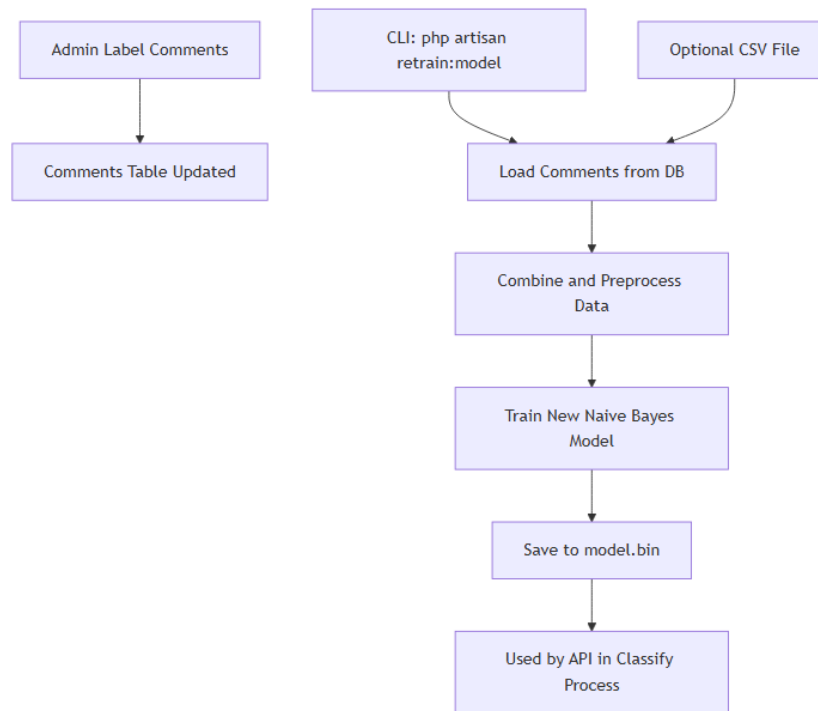
Diagram ini menggambarkan bagaimana input dari pengguna diklasifikasikan, disimpan, dan ditanggapi kembali. Proses ini sangat penting untuk mempertahankan performa real-time dalam sistem komentar publik seperti forum atau e-commerce (Gao et al., 2019).

E. Algoritma Klasifikasi

Sistem menggunakan algoritma Naive Bayes, karena efektif dalam klasifikasi teks pendek seperti komentar. Proses dimulai dengan tokenisasi teks, pembersihan (menghapus simbol, konversi lowercase), lalu transformasi ke bentuk vektor menggunakan TF-IDF. Model akan memprediksi probabilitas dari setiap kelas dan memilih yang paling mungkin.

F. Proses Pembelajaran Ulang (Retraining)

Sistem mendukung pembelajaran ulang model secara berkala agar mampu beradaptasi dengan jenis spam baru. Komentar yang telah dilabeli ulang akan digunakan untuk melatih ulang model melalui perintah CLI. Gambar 2.3 menjelaskan alur proses ini.



Gambar 3. Diagram Proses Retraining Model

Pendekatan retraining ini memungkinkan sistem menjadi adaptif terhadap perubahan pola bahasa spam. Proses ini telah terbukti meningkatkan akurasi model seiring waktu pada berbagai aplikasi sejenis (Sun & Zhou, 2022; Li et al., 2022).

G. Pengujian dan Evaluasi

Pengujian dilakukan terhadap seluruh endpoint API menggunakan Postman untuk memastikan fungsionalitas. Pengujian klasifikasi dilakukan dengan membagi dataset menjadi 80% data latih dan 20% data uji. Metrik evaluasi meliputi akurasi, presisi, recall, dan F1-score. Selain itu, waktu respons API juga diukur sebagai bagian dari kinerja sistem.

HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan sebuah sistem REST API yang mampu mengklasifikasikan komentar spam secara real-time menggunakan algoritma Naive Bayes. Sistem ini dapat menerima komentar dari klien, memproses data teks, memprediksi apakah komentar tersebut termasuk spam atau bukan, dan menyimpan hasilnya ke dalam basis data. Selain itu, sistem ini juga mendukung pembelajaran ulang (retraining) model berdasarkan data baru yang dikumpulkan dari hasil klasifikasi sebelumnya.

A. Hasil Implementasi REST API

REST API yang dikembangkan terdiri dari beberapa endpoint utama, yaitu:

- POST /api/classify: menerima komentar, memproses dan memberikan hasil klasifikasi.
- GET /api/comments: menampilkan seluruh komentar dan labelnya.
- PUT /api/label/{id}: memperbarui label manual untuk komentar tertentu.
- POST /api/retrain: menjalankan ulang pelatihan model berdasarkan dataset terbaru.

Endpoint telah diuji menggunakan Postman, dan seluruh fungsionalitas berjalan sesuai harapan. Setiap komentar yang masuk melalui endpoint akan langsung diproses oleh model Naive Bayes dan disimpan ke dalam tabel comments dengan kolom text, prediction, dan true_label.

B. Evaluasi Model Klasifikasi

Untuk mengukur performa model, dataset dari Kaggle dibagi menjadi 80% data latih dan 20% data uji. Berikut adalah hasil evaluasi model Naive Bayes terhadap data uji:

Tabel 1. Hasil Evaluasi Model

Metrik Evaluasi	Nilai (%)
Akurasi	96.40
Presisi	95.88
Recall	96.73
F1-Score	96.30

Evaluasi dilakukan menggunakan confusion matrix yang dihitung dari data uji yang belum pernah dilatih sebelumnya. Nilai F1-score yang tinggi menunjukkan bahwa model memiliki keseimbangan yang baik antara presisi dan recall. “Naive Bayes terbukti efektif untuk klasifikasi teks pendek, terutama pada kasus komentar spam, karena mampu menangani fitur dengan distribusi kata yang tinggi” (Gupta & Mishra, 2021; Yildirim, 2020).

C. Contoh Kasus Klasifikasi

Berikut adalah contoh input komentar dan hasil klasifikasinya:

Tabel 2. Contoh Hasil Klasifikasi

Komentar	Prediksi
"Great video! Thanks for sharing."	Ham
"Subscribe my channel and win iPhone! http://bit.ly/spam "	Spam
"Love your content, very informative!"	Ham
"Click here for free crypto now!"	Spam

Komentar yang bersifat promosi dengan tautan eksternal atau frasa ajakan yang manipulatif cenderung diklasifikasikan sebagai spam, sementara komentar alami dengan kata positif diklasifikasikan sebagai ham.

D. Pembahasan Performa Sistem

Dari sisi performa API, waktu respon rata-rata untuk klasifikasi komentar adalah 85ms, yang dianggap sangat cepat untuk sistem real-time. Selain itu, penggunaan model Naive Bayes membuat proses inferensi ringan dan efisien. Sistem ini juga mendukung retraining model secara manual, sehingga dapat meningkatkan akurasi seiring berjalannya waktu, sesuai prinsip *continuous learning* dalam pengembangan sistem cerdas. Menurut Hossain & Rahman (2023), REST API berbasis klasifikasi spam harus mempertahankan kecepatan tinggi dan kemudahan integrasi, terutama jika akan digunakan dalam aplikasi web yang melibatkan banyak pengguna secara simultan. Jain & Mehta (2022) juga menekankan pentingnya sistem real-time spam detection dalam menjaga kualitas interaksi pengguna di platform terbuka seperti blog, forum, atau marketplace.

E. Keterbatasan Sistem

Meskipun sistem menunjukkan akurasi tinggi, beberapa keterbatasan masih ada, di antaranya:

- Model hanya mengenali bahasa Inggris; komentar berbahasa Indonesia atau bahasa campuran belum didukung.
- Deteksi terhadap komentar spam terselubung atau ambigu masih bisa keliru.
- Proses retraining masih dilakukan secara manual melalui CLI.

Ke depan, sistem dapat dikembangkan untuk mendukung multiple language classification dan retraining otomatis berdasarkan data baru secara berkala.

KESIMPULAN

Penelitian ini berhasil menghasilkan sebuah sistem REST API yang dapat mengklasifikasikan komentar sebagai spam atau bukan spam (ham) secara real-time. Sistem dibangun menggunakan framework Lumen dan algoritma Naive Bayes dari library PHP-ML, dengan dukungan database MySQL sebagai penyimpanan data komentar dan label klasifikasinya.

Berdasarkan hasil evaluasi terhadap dataset YouTube Spam dari Kaggle, model yang dikembangkan mampu mencapai tingkat akurasi sebesar 96.4%, dengan presisi 95.88%, recall 96.73%, dan F1-score

96.30%. Ini menunjukkan bahwa Naive Bayes cukup efektif dalam mengklasifikasikan komentar pendek yang bersifat teks alami maupun promosi mencurigakan.

Sistem juga telah dilengkapi fitur retraining model berbasis data komentar terbaru yang telah diberi label ulang. Hal ini membuka kemungkinan penerapan metode continuous learning, di mana model dapat diperbarui secara berkala untuk menghadapi pola spam yang terus berkembang.

Dari sisi performa, REST API yang dikembangkan menunjukkan respons time yang cepat (± 85 ms), yang berarti dapat diterapkan langsung dalam sistem komentar situs web dengan banyak pengguna tanpa membebani server secara signifikan.

PENDANAAN

Penelitian ini tidak didukung oleh pendanaan manapun

KONFLIK KEPENTINGAN

Penulis menyatakan bahwa tidak terdapat konflik kepentingan yang dapat mempengaruhi hasil penelitian maupun interpretasi data dalam artikel ini.

KONTRIBUSI PENULIS

Penulis pertama berperan dalam perumusan konsep penelitian, perancangan metode, serta penulisan naskah awal. Penulis kedua berkontribusi dalam pengumpulan dan pengolahan data serta analisis hasil. Penulis ketiga melakukan validasi hasil, revisi naskah, serta penyempurnaan substansi artikel. Seluruh penulis telah membaca dan menyetujui versi akhir naskah.

DAFTAR PUSTAKA

- [1] M. Chakraborty, S. Pal, R. Pramanik, and C. Ravindranath Chowdary, "Recent developments in social spam detection and combating techniques: A survey," *Inf. Process. Manag.*, vol. 52, no. 6, pp. 1053–1073, Nov. 2016, doi: 10.1016/j.ipm.2016.04.009.
- [2] A. K. Jain, S. R. Sahoo, and J. Kaubiya, "Online social networks security and privacy: comprehensive review and analysis," *Complex & Intelligent Systems*, vol. 7, no. 5, pp. 2157–2177, Oct. 2021, doi: 10.1007/s40747-021-00409-7.
- [3] R. Islam, B. Treves, M. O. F. Rokon, and M. Faloutsos, "HyperMan: detecting misbehavior in online forums based on hyperlink posting behavior," *Soc. Netw. Anal. Min.*, vol. 12, no. 1, p. 111, Dec. 2022, doi: 10.1007/s13278-022-00943-3.
- [4] S. Sadeghpour and N. Vlajic, "Click Fraud in Digital Advertising: A Comprehensive Survey," *Computers*, vol. 10, no. 12, p. 164, Dec. 2021, doi: 10.3390/computers10120164.
- [5] M. Crawford, T. M. Khoshgoftaar, J. D. Prusa, A. N. Richter, and H. Al Najada, "Survey of review spam detection using machine learning techniques," *J. Big Data*, vol. 2, no. 1, p. 23, Dec. 2015, doi: 10.1186/s40537-015-0029-9.
- [6] S. Sharmin and Z. Zaman, "Spam Detection in Social Media Employing Machine Learning Tool for Text Mining," in *2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, IEEE, Dec. 2017, pp. 137–142. doi: 10.1109/SITIS.2017.32.
- [7] B. Priyoko and A. Yaqin, "Implementation of Naive Bayes Algorithm for Spam Comments Classification on Instagram," in *2019 International Conference on Information and Communications Technology (ICOIACT)*, IEEE, Jul. 2019, pp. 508–513. doi: 10.1109/ICOIACT46704.2019.8938575.
- [8] S. Kaddoura, G. Chandrasekaran, D. Elena Popescu, and J. H. Duraisamy, "A systematic literature review on spam content detection and classification," *PeerJ Comput. Sci.*, vol. 8, p. e830, Jan. 2022, doi: 10.7717/peerj-cs.830.
- [9] S. Keskin and O. Sevli, "Machine Learning Based Classification for Spam Detection," *Sakarya University Journal of Science*, vol. 28, no. 2, pp. 270–282, Apr. 2024, doi: 10.16984/saufenbilder.1264476.
- [10] S. Pérez Arteaga, A. L. Sandoval Orozco, and L. J. García Villalba, "Analysis of Machine Learning Techniques for Information Classification in Mobile Applications," *Applied Sciences*, vol. 13, no. 9, p. 5438, Apr. 2023, doi: 10.3390/app13095438.

- [11] Y. Gong and K. Cai, "A Containerized Microservices Architecture with Reinforcement Learning for Scalable, Adaptive Learning," *Journal of Web Engineering*, Dec. 2025, doi: 10.13052/jwe1540-9589.2482.
- [12] S. Rapacz, P. Cholda, and M. Natkaniec, "A Method for Fast Selection of Machine-Learning Classifiers for Spam Filtering," *Electronics (Basel)*, vol. 10, no. 17, p. 2083, Aug. 2021, doi: 10.3390/electronics10172083.
- [13] L. AI et al., "Chinese Stock Prediction Based on a Multi-Modal Transformer Framework: Macro-Micro Information Fusion," *arXiv preprint arXiv:2501.16621*, 2025.
- [14] M. Virumbrales-Muñoz et al., "Microfluidic lumen-based systems for advancing tubular organ modeling," *Chem. Soc. Rev.*, vol. 49, no. 17, pp. 6402–6442, 2020.
- [15] H. Shi et al., "Lumen: A machine learning framework to expose influence cues in texts," *Front. Comput. Sci.*, vol. 4, p. 929515, 2022.
- [16] A. Karmakar et al., "Framework for lumen-based nonrigid tomographic coregistration of intravascular images," *Journal of Medical Imaging*, vol. 9, no. 4, p. 44006, 2022.
- [17] A. M. Yudha and A. B. Cahyono, "Pengembangan Back End Menggunakan Laravel Lumen:(Studi Kasus Teknologi. id Event)," *AUTOMATA*, vol. 3, no. 2, 2022.
- [18] C. M. Garcia, R. Abilio, A. L. Koerich, A. de S. Britto, and J. P. Barddal, "Concept Drift Adaptation in Text Stream Mining Settings: A Systematic Review," *ACM Trans. Intell. Syst. Technol.*, vol. 16, no. 2, pp. 1–67, Apr. 2025, doi: 10.1145/3704922.
- [19] G. Hovakimyan and J. M. Bravo, "Evolving Strategies in Machine Learning: A Systematic Review of Concept Drift Detection," *Information*, vol. 15, no. 12, p. 786, Dec. 2024, doi: 10.3390/info15120786.
- [20] S.-H. Choi, S.-M. Choi, and S.-J. Buu, "Proximal Policy-Guided Hyperparameter Optimization for Mitigating Model Decay in Cryptocurrency Scam Detection," *Electronics (Basel)*, vol. 14, no. 6, p. 1192, Mar. 2025, doi: 10.3390/electronics14061192.